MEASURES OF DISPERSION FOR GROUPED DATA

BY : RRAEZZZ

❑ CONSTRUCT FREQUENCY POLYGON

Size of class interval
$= \left(\frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}\right)$

Lower limit is the smallest value and upper limit is the largest value of a class.

Midpoint $= \left(\frac{\text{Lower limit} + \text{Upper limit}}{2}\right)$

Lower boundary
$= \left(\frac{\text{Upper limit of the class before it} + \text{Lower limit of the class}}{2}\right)$

Upper boundary
$= \left(\frac{\text{Upper limit of the class} + \text{Lower limit of the class after it}}{2}\right)$

The data on the right shows the heights, to the nearest cm, of a group of Form 5 pupils.

153	168	163	157
158	161	165	162
145	150	158	156
166	163	152	155
158	173	148	164

(a) Determine the class intervals for the data, if the number of classes required is 6.

(b) Construct a frequency table based on the information in (a). Hence, complete the frequency table with the lower limit, upper limit, midpoint, lower boundary and upper boundary.

(a) The largest data is 173 and the smallest data is 145.
If the number of classes is 6, then the size of each class interval

$= \frac{173 - 145}{6}$
$= 4.7 \approx 5$

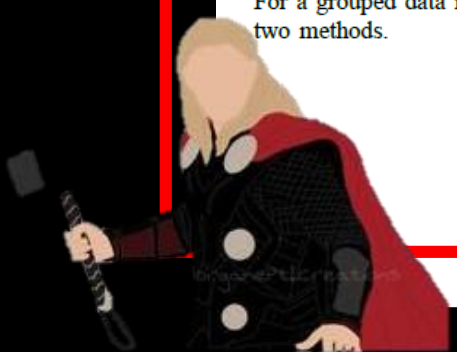
Size of class interval
$= \left(\frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}\right)$

Therefore, the class intervals are 145 – 149, 150 – 154, 155 – 159, 160 – 164, 165 – 169 and 170 – 174.

(b)

Height (cm)	Frequency	Lower limit	Upper limit	Midpoint	Lower boundary	Upper boundary
145 – 149	2	145	149	147	144.5	149.5
150 – 154	3	150	154	152	149.5	154.5
155 – 159	6	155	159	157	154.5	159.5
160 – 164	5	160	164	162	159.5	164.5
165 – 169	3	165	169	167	164.5	169.5
170 – 174	1	170	174	172	169.5	174.5

For a grouped data in uniform class intervals, the size of class interval can be calculated using two methods.



BY : RRAEZZZ

7.1 DISPERSION

- The cumulative frequency of a data can also be obtained from a frequency table
- The cumulative frequency of a class interval is the sum of the frequency of the class and the otal frequency of the classes before it
- This gives an ascending cumulative frequency

Method 1: The difference between the lower limits or the upper limits of two consecutive classes.

Size of class interval of the first two classes
= 150 − 145 ← Lower limit of class 145 – 149
= 5 ↑ Lower limit of class 150 – 154
or
= 154 − 149 ← Upper limit of class 145 – 149
= 5 ↑ Upper limit of class 150 – 154

Method 2: The difference between the upper boundary and the lower boundary of a class interval.

Size of class interval of the first class
= 149.5 − 144.5 ← Lower boundary of class 145 – 149
= 5 ↑ Upper boundary of class 145 – 149

Construct a cumulative frequency table from the frequency table below.

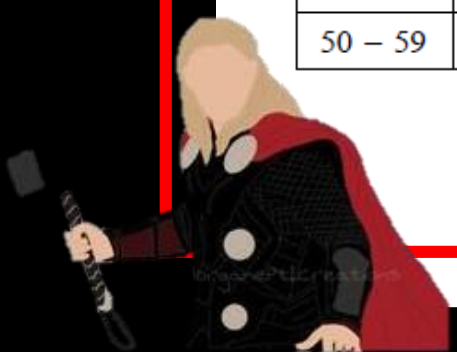
Age	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59
Frequency	4	5	8	7	3

Solution:

Age	Frequency	Cumulative frequency
10 – 19	4	4
20 – 29	5	9
30 – 39	8	17
40 – 49	7	24
50 – 59	3	27

(+ 4 → 9; + 8 → 17; + 7 → 24; + 3 → 27)

This value of 17 means there are 17 people aged 39 years old and below

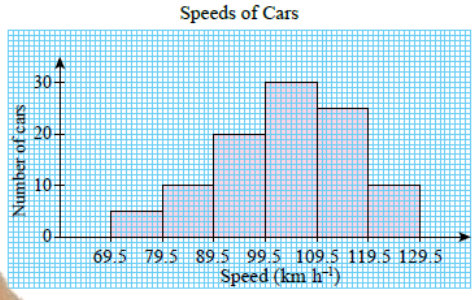


BY : RRAEZZZ

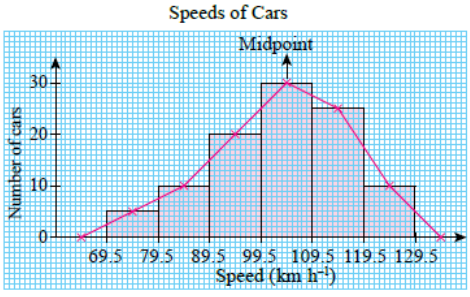
7.1 DISPERSION

- Histogram is a graphical representation in which the data is grouped into ranges by using contiguous bars.
- The height of the bar in histogram represents the frequency of a class
- Steps for constructing a histogram
  1) Find the lower boundary and upper boundary of each class interval
  2) Choose an appropriate scale on the vertical axis. Represent the frequencies on the vertical axis and the class boundaries on the horizontal axis
  3) Draw bars that represent each class where the width is equal to the size of the class and the height is a proportionate to the frequency
- Frequency polygon is a graph that displays a grouped data by using straight lines that connect midpoints of the classes which lie at the upper end of each bar in a histogram
- Steps for constructing a frequency polygon
  1) Mark the midpoints of each class on top of each bar
  2) Mark the midpoints before the first class and after the last class with zero frequency
  3) Draw straight lines by connecting the adjacent midpoints

Histogram:



Frequency polygon:



BY : RRAEZZZ

7.1 DISPERSION

❑ Steps for constructing a frequency polygon without constructing a histogram

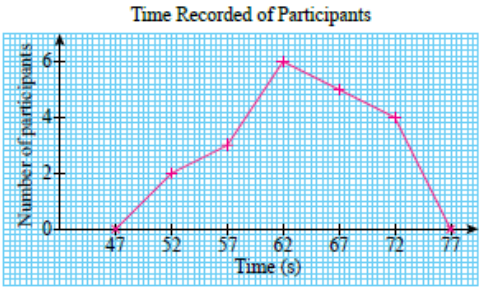
1) Add one class interval before the first class and after the last class with zero frequency
2) Find the midpoint of each class interval
3) Choose an appropriate scale on the vertical axis. Represent the frequencies on the vertical axis and the midpoints on the horizontal axis
4) Mark the midpoint with the corresponding frequency
5) Connect each midpoint with a straight line

The frequency table below shows the time in seconds, recorded by 20 participants in a qualifying round of a swimming competition. Represent the data with a frequency polygon by using a scale of 2 cm to 5 seconds on the horizontal axis and 2 cm to 2 participants on the vertical axis.

Time recorded (s)	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74
Number of participants	2	3	6	5	4

Solution:

Time recorded (s)	Number of participants	Midpoint
45 – 49	0	47
50 – 54	2	52
55 – 59	3	57
60 – 64	6	62
65 – 69	5	67
70 – 74	4	72
75 – 79	0	77

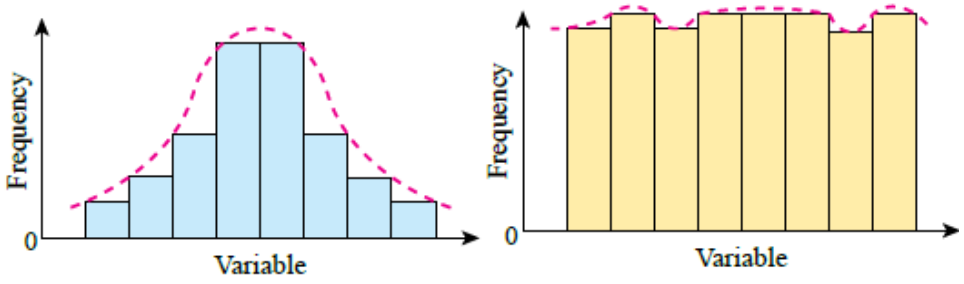


Add a class interval with zero frequency before the first class and after the last class

BY : RRAEZZZ

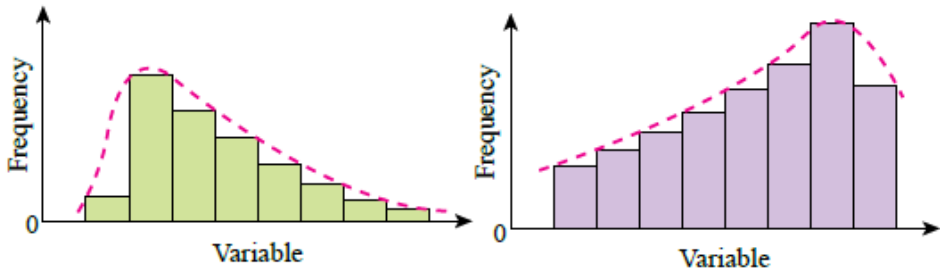
7.1 DISPERSION

❑ Symmetric Histogram



Bell-shaped **Uniform-shaped**

❑ Skewed Histogram



Right-skewed **Left-skewed**

❑ Steps for constructing an ogive

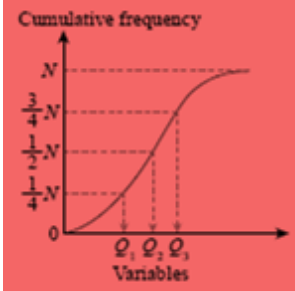
1) Add one class before the first class with zero frequency. Find the upper boundary and the cumulative frequency for each class
2) Choose an appropriate scale on the vertical axis to represent the cumulative frequencies and the horizontal axis to represent the upper boundaries
3) Plot the cumulative frequency with the corresponding upper boundary
4) Draw a smooth curve passing through all the points

BY : RRAEZZZ

7.1 DISPERSION

Quartile

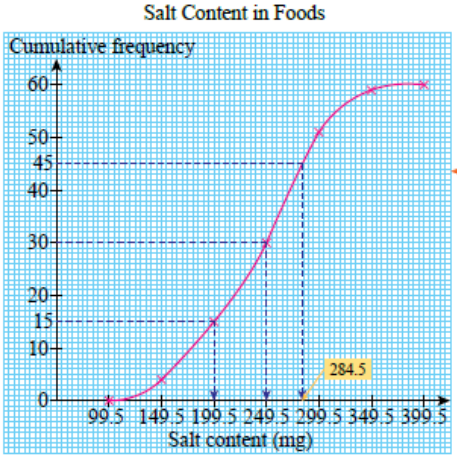
For a grouped data with number of data N, the quartiles can be determined from the ogive. Q_1, Q_2 and Q_3 are the values that correspond to the cumulative frequency $\frac{N}{4}$, $\frac{N}{2}$ and $\frac{3N}{4}$ respectively.



The first quartile position, Q_1
$= \frac{1}{4} \times$ total frequency, N

The median position, Q_2
$= \frac{1}{2} \times$ total frequency, N

The third quartile position, Q_3
$= \frac{3}{4} \times$ total frequency, N



Steps to determine the quartiles:

1. Number of data, $N = 60$, therefore $\frac{N}{4} = 15$, $\frac{N}{2} = 30$ and $\frac{3N}{4} = 45$.
2. Draw a horizontal line from the axis of cumulative frequency at 15 until it intersects the ogive.
3. From the intersection point in step 2, draw the vertical line down until it meets the axis of salt content at the horizontal axis.
4. The value of the salt content obtained is the value of Q_1.
5. Repeat steps 2 to 4 for the values of 30 and 45 to obtain the values of Q_2 and Q_3.

Percentile

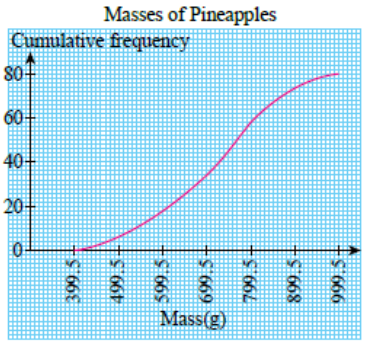
We can analyse a large data more easily and effectively when we divide the data into small parts which is known as percentile. A percentile is a value that divides a set of data into 100 equal parts and is represented by P_1, P_2, P_3, ..., P_{99}.

BY : RRAEZZZ

Range And interquartile range

Pak Hamidi had recorded the mass of pineapples that he harvested from his farm. The following frequency table and ogive show the data that he obtained. Determine the range and interquartile range for the data.

Mass (g)	Number of pineapples
400 – 499	6
500 – 599	12
600 – 699	16
700 – 799	24
800 – 899	14
900 – 999	8



MEMORY BOX

Interquartile range
$= Q_3 - Q_1$

Smart TIPS

Interquartile range of a set of grouped data can be determined from ogive by finding Q_1 and Q_3 first.

Solution:

Range = midpoint of the highest class – midpoint of the lowest class

$$= \frac{900 + 999}{2} - \frac{400 + 499}{2}$$

$$= 949.5 - 449.5$$

$$= 500 \text{ g}$$

Difference between the heaviest pineapple and the lightest pineapple is 500 g.

From the ogive,

the position of Q_1:

$$\frac{1}{4} \times 80 = 20$$

$Q_1 = 614.5$

the position of Q_3:

$$\frac{3}{4} \times 80 = 60$$

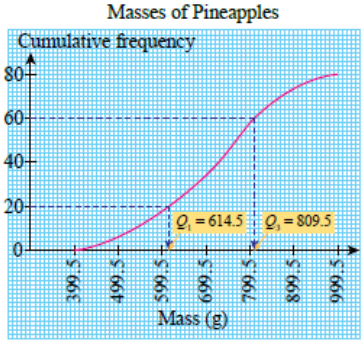
$Q_3 = 809.5$

Therefore, the interquartile range

$= 809.5 - 614.5$

$= 195$ g

Difference between the heaviest pineapple and the lightest pineapple that lies in the middle 50% of the distribution is 195 g.



Variance and Standard Deviation

Variance, $\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2$

Standard deviation, $\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$

where

x = midpoint of the class interval

f = frequency

$\bar{x}$ = mean of the data

BY : RRAEZZZ

❑ Variance And Standard Deviation

The frequency table below shows the volumes of water to the nearest litres, used daily by a group of families in a housing area. Calculate the variance and standard deviation of the data.

Volume of water (ℓ)	150 – 159	160 – 169	170 – 179	180 – 189	190 – 199	200 – 209
Number of families	8	12	15	24	20	16

Solution:

Volume of water (ℓ)	Frequency, f	Midpoint, x	fx	x^2	fx^2
150 – 159	8	154.5	1 236	23 870.25	190 962
160 – 169	12	164.5	1 974	27 060.25	324 723
170 – 179	15	174.5	2 617.5	30 450.25	456 753.75
180 – 189	24	184.5	4 428	34 040.25	816 966
190 – 199	20	194.5	3 890	37 830.25	756 605
200 – 209	16	204.5	3 272	41 820.25	669 124
	$\sum f = 95$		$\sum fx = 17\ 417.5$		$\sum fx^2 = 3\ 215\ 133.75$

$$\text{Mean, } \bar{x} = \frac{\sum fx}{\sum f}$$
$$= \frac{17\ 417.5}{95}$$
$$= 183.34\ \ell$$

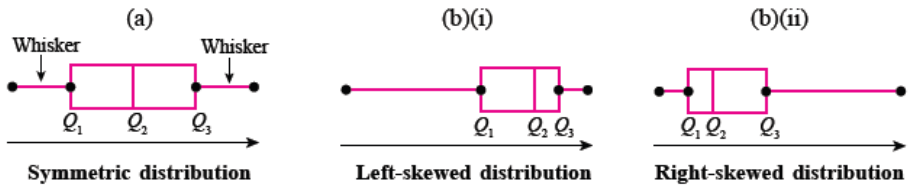
$$\text{Variance, } \sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2$$
$$= \frac{3\ 215\ 133.75}{95} - \left(\frac{17\ 417.5}{95}\right)^2$$
$$= 229.1856$$
$$= 229.19\ \ell^2 \text{ (correct to 2 decimal places)}$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$
$$= \sqrt{229.1855956}$$
$$= 15.1389$$
$$= 15.14\ \ell \text{ (correct to 2 decimal places)}$$

Checking Answer

1. Press MODE MODE
   Display: SD REG BASE / 1 2 3
   Choose 1
2. Enter (midpoint), press SHIFT ↑ (frequency) M+, and repeat for the subsequent values.
3. Press AC SHIFT 2
   Display: $\bar{x}$ $x\sigma n$ $x\sigma n-1$ / 1 2 3
   Press 1 for mean:
   Display 183.3421053
   Press 2 for standard deviation:
   Display 15.13887696

❑ Construct and interpret a box plot for a set of grouped data



BY : RRAEZZZ

❑ Construct and interpret a box plot for a set of grouped data

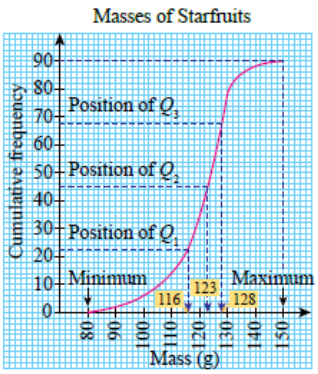
(a) The median lies in the middle of the box and the whiskers are about the same length on both sides of the box.
(b) The median cuts the box into two different sizes.
  (i) If the left side of the box is longer, then the data distribution is left-skewed.
  (ii) If the right side of the box is longer, then the data distribution is right-skewed.

Left whisker and right whisker represent the score outside of the median. If the box is divided into the same size but the left whisker is longer than the right whisker, then the data distribution is left-skewed, and vice versa.

Example 11

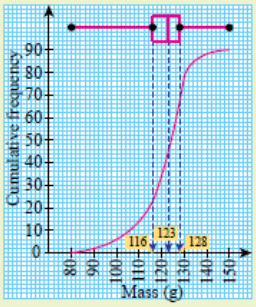
The ogive on the right shows the masses in g, of 90 starfruits.

(a) Construct a box plot based on the ogive.
(b) Hence, state the distribution shape of the data.

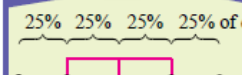


Info Bulletin

Ogive and box plot on the same graph:



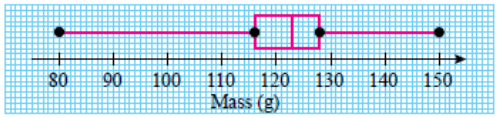
Info Bulletin



Solution:

(a) From the ogive:
- Minimum value = 80
- Maximum value = 150
- Position of Q_1: $\frac{1}{4} \times 90 = 22.5$
  $Q_1 = 116$
- Position of Q_2: $\frac{1}{2} \times 90 = 45$
  $Q_2 = 123$
- Position of Q_3: $\frac{3}{4} \times 90 = 67.5$
  $Q_3 = 128$

Box plot:



(b) The distribution of the data is skewed to the left because the left side of the box plot is longer than the right side of the box plot.

BY : RRAEZZZ